

Nuclear magnetic resonance-based determination of dioxygen binding sites in protein cavities

Ryo Kitahara ^{1,*} Shun Sakuraba,² Tomoshi Kameda,³ Sanshiro Okuda,⁴ Mengjun Xue,⁵ and Frans A.A. Mulder ⁵

¹Department of Pharmaceutical Sciences, College of Pharmaceutical Sciences, Ritsumeikan University, Kusatsu, 525-8577, Japan

²Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, 277-8561, Japan

³Computational Omics Research Team, Artificial Intelligence Research Center, Advanced Industrial Science and Technology, Koto, Tokyo, 135-0064, Japan

⁴Graduate School of Life Sciences, Ritsumeikan University, Kusatsu, 525-8577, Japan

⁵Department of Chemistry and Interdisciplinary Nanoscience Center (iNANO), University of Aarhus, Aarhus C, DK-8000, Denmark

Received 5 September 2017; Accepted 20 December 2017

DOI: 10.1002/pro.3371

Published online 22 December 2017 proteinscience.org

Abstract: The location and ligand accessibility of internal cavities in cysteine-free wild-type T4 lysozyme was investigated using O₂ gas-pressure NMR spectroscopy and molecular dynamics (MD) simulation. Upon increasing the concentration of dissolved O₂ in solvent to 8.9 mM, O₂-induced paramagnetic relaxation enhancements (PREs) to the backbone amide and side chain methyl protons were observed, specifically around two cavities in the C-terminal domain. To determine the number of O₂ binding sites and their atomic coordinates from the 1/*r*⁶ distance dependence of the PREs, we established an analytical procedure using Akaike's Information Criterion, in combination with a grid-search. Two O₂-accessible sites were identified in internal cavities: One site was consistent with the xenon-binding site in the protein in crystal, and the other site was established to be a novel ligand-binding site. MD simulations performed at 10 and 100 mM O₂ revealed dioxygen ingress and egress as well as rotational and translational motions of O₂ in the cavities. It is therefore suggested that conformational fluctuations within the ground-state ensemble transiently develop channels for O₂ association with the internal protein cavities.

Keywords: T4 lysozyme; gas-pressure NMR; molecular dynamics simulation; paramagnetic relaxation enhancement; oxygen

Abbreviations: NMR, nuclear magnetic resonance; PRE, paramagnetic relaxation enhancement; MD, molecular dynamics; HSQC, heteronuclear single-quantum coherence; AIC, Akaike Information Criterion; CMA-ES, covariance matrix adaptation evolution strategy.

Additional Supporting Information may be found in the online version of this article.

Funding: This work was supported by research grants from the Institute for Chemical Fibers, Japan and Ritsumeikan University to R.K.

Notes: The authors declare no competing financial interest.

Grant sponsor: the Institute for Chemical Fibers, Japan; Grant sponsor: Ritsumeikan University.

*Correspondence to: [Ryo Kitahara, College of Pharmaceutical Sciences, Ritsumeikan University, Kusatsu 525-8577, Japan]. E-mail: ryo@ph.ritsumei.ac.jp

Introduction

Internal protein cavities can accommodate small molecules such as water and organic compounds.^{1,2} Exploring internal protein cavities and pockets helps improve our understanding of protein stability, conformational fluctuation, and ligand binding.^{3,4} For example, cavity enlargement can enhance protein conformational fluctuation and decrease thermal stability,^{4,5} while ligand binding to internal cavities as well as cavity filling by amino acid substitutions can increase protein thermal stability.^{6–8}

Cysteine-free wild-type T4 lysozyme, also called pseudo-wild-type T4 lysozyme (WT*), has similar structural and functional properties to the wild-type T4 lysozyme despite C54T and C97A mutations.⁹ The

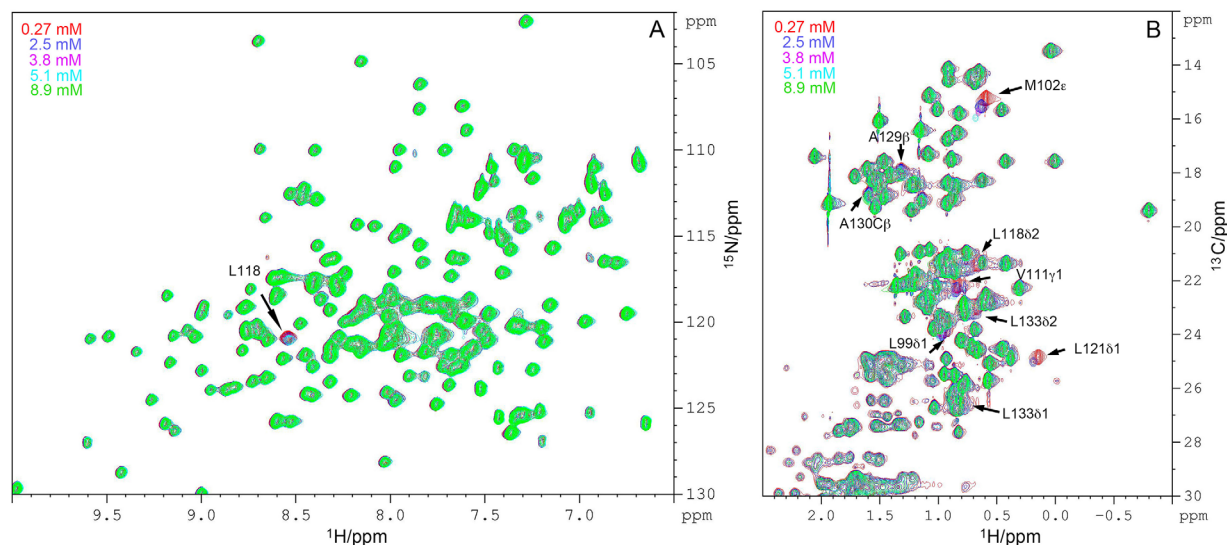


Figure 1. O₂-induced NMR spectral changes. ¹H/¹⁵N HSQC (A) and ¹H/¹³C constant time HSQC (B) spectra of ¹³C/¹⁵N double-labeled WT* T4 lysozyme at 298 K in association with O₂ concentrations ranging from 0.27 to 8.9 mM. Positive and negative cross-peaks are represented by the same color. Amide and methyl groups showing significant chemical shift changes or a loss of peak intensity are indicated by arrows (→).

WT* C-terminal domain has two small cavities (cavities 2 and 3) and one somewhat larger cavity (cavity 4, ~48 Å³), while the N-terminal domain has one small cavity (cavity 1). As determined by X-ray crystallography, electron density indicative of water molecules was observed in cavity 1, but not in cavities 2, 3, and 4.⁹ Moreover, the pressure-dependent binding of noble gases such as xenon, krypton, and argon to WT* and its mutants was only observed in the larger cavity (i.e., cavity 4) of the protein in crystal.¹⁰ The affinity of noble gases seems to be understood in terms of cavity size and ligand polarizability. Native state hydrogen exchange nuclear magnetic resonance (NMR), with a denaturant concentration gradient, showed that the extrapolated free energy for unfolding the WT* C-terminal domain was significantly higher (62 ± 8 kJ/mol) than that for the N-terminal domain (35 ± 7 kJ/mol) in the absence of denaturant. In contrast, the free energy of hydrogen exchange for the C- and N-terminal domains were similar (~20–30 kJ/mol) under identical conditions.¹¹ These results indicate that the WT* C-terminal domain is highly stable against unfolding, yet has high local fluctuations that enable hydrogen exchange. These local fluctuations within the ground-state ensemble may facilitate binding of small ligands. However, knowledge of ligand binding to WT* and its cavity-designed mutants in solution remains limited.

NMR provides a high-resolution approach to investigate the structure and dynamics of proteins in solution. Several studies have shown that NMR spectroscopy in combination with dioxygen (O₂) gas-pressure is a powerful tool for investigating the location of internal protein cavities and pockets in solution and their accessibility to ligands, owing to the paramagnetic

properties of O₂.^{12–18} In previous studies, methods to analyze O₂-binding sites in proteins in solution were reported, using the O₂-induced enhancement of backbone amide proton spin-lattice relaxation.^{12,15,16} We found that O₂ bound selectively to the hydrophobic, rather than hydrophilic (water-containing) cavities in the cavity-enlarged T4 lysozyme mutant L99A.¹² In the outer surface protein A (OspA), O₂ was also found in association with the large hydrophobic cavity (~200 Å).¹⁶ Although O₂-induced changes in the ¹H/¹³C HSQC spectra for WT* were previously reported,¹² the details thereof have not yet been reported.

Here, we report on the O₂-binding sites in WT* and their accessibility to ligands using O₂ gas-pressure NMR spectroscopy and molecular dynamics (MD) simulation. In addition, we established an analytical procedure using the Akaike Information Criterion (AIC),¹⁹ which is one of the popular indices in statistics for facilitating statistical model identification, and a grid-search to determine the number of O₂ binding sites and their atomic coordinates from the 1/*r*⁶ distance dependence of O₂-induced paramagnetic relaxation enhancement (PRE).

Results and Discussion

O₂-induced spectral changes

The association of O₂ with WT* was studied by solution NMR spectroscopy at an O₂ concentration range of 0.27–8.9 mM. All measurements were initiated >19 h after changing the gas pressure, because 19 h was required to reach a new gas dissolution equilibrium in the NMR tube when studying the L99A mutant protein in which larger NMR spectrum changes were observed.¹² Figure 1 shows ¹H/¹⁵N

HSQC and $^1\text{H}/^{13}\text{C}$ CT-HSQC spectra of the protein at different O_2 concentrations. Significant chemical shift changes and peak broadening were only observed for the amide group of L118. Large chemical shift changes and/or peak broadening were observed for the methyl groups M102 ϵ , V111 γ 1, L118 δ 2, L121 δ 1, A129 β , A130 β , L133 δ 1, and L133 δ 2. These spectral changes were O_2 concentration dependent and fully reversible. Because these spectral changes were not observed when N_2 was increased to 7 bar (corresponding to 4.6 mM), the chemical shift and peak-width changes are considered to originate from the paramagnetic properties of O_2 . These results indicate that O_2 molecules penetrate the hydrophobic interior of the protein, thereby causing the observed changes. Indeed, similar O_2 -induced chemical shifts and peak broadening were observed in the cavity enlarged mutant L99A of the protein¹² and OspA,¹⁶ both of which contain interior hydrophobic cavities.

O_2 -induced paramagnetic relaxation enhancements

Paramagnetic triplet O_2 has two unpaired electrons. In order to investigate O_2 association with WT*, O_2 -induced PREs were measured in the presence of 4 (5.1 mM) and 7 bar (8.9 mM) O_2 . Figure 2(A) shows longitudinal relaxation rate constants R_1 for amide protons, according to the residue number, at 7 bar of O_2 or N_2 . PREs, defined as R_1 differences between O_2 and N_2 , are shown in Figure 2(B). O_2 -induced increases in R_1 of $>2.0 \text{ s}^{-1}$ were observed at several residue sites among residues 110–155, while increases in R_1 for the remaining residues were about 1.0 s^{-1} . To visualize the O_2 binding sites in proteins, amide groups showing large PREs were mapped on the structural model of WT*, as shown in Figure 3(A). Except for residue 6 and 19, all amide groups showing increases in R_1 of $>2.0 \text{ s}^{-1}$ at 7 bar of O_2 were located around hydrophobic cavities 3 and 4 in the C-terminal domain.

The R_1 for methyl protons are shown in Figure 4. Methyl groups showing large ^1H longitudinal relaxation enhancements ($4 \text{ s}^{-1} < \Delta R_1 < 10 \text{ s}^{-1}$, orange; $\Delta R_1 \geq 10 \text{ s}^{-1}$ or marked signal loss, red) at 8.9 mM O_2 were mapped in Figure 3(B). As with the amide protons, methyl protons exhibiting large relaxation enhancements were located around the hydrophobic cavities. These results strongly indicate the association of O_2 with hydrophobic cavities 3 and 4.

Analysis of O_2 -binding sites

PREs originate from unpaired electrons of the paramagnetic triplet O_2 . We previously proposed a method to analyze O_2 -accessible sites in proteins based on the $1/r^6$ distance-dependent PRE.¹⁶ ^1H longitudinal relaxation enhancements, ΔR_1 , was

expressed by the $1/r^6$ distance-dependent PRE contributions from each O_2 -binding site. When the number of O_2 -binding sites is n , ΔR_1 can be expressed using the following Eq. (1):

$$\Delta R_1 = C_0 + \sum_{i=1}^n C_i \left(\frac{1}{r_i} \right)^6 \quad (1)$$

where r_i are the distances (\AA) to the O_2 -binding site i , C_i is constant that is proportional to the PRE contributions from each O_2 -binding site, and C_0 is the background PRE contribution from O_2 dissolved in the solvent. Because the distances between the O_2 -binding site and the nucleus of interest are calculated from their atomic coordinates, Eq. (1) can be expressed by Eq. (2):

$$\Delta R_1 = C_0 + \sum_{i=1}^n C_i \left((x_a - X_i)^2 + (y_a - Y_i)^2 + (z_a - Z_i)^2 \right)^{-3} \quad (2)$$

where x_a , y_a , and z_a are the atomic coordinates of each amide proton, and X_i , Y_i , and Z_i are the atomic coordinates of bound O_2 in the three-dimensional Cartesian coordinate system. The observed ΔR_1 values were globally fitted to Eq. (2) and C_0 , C_i , X_i , Y_i , and Z_i were obtained as fitting parameters.

In the previous study, the initial parameters for nonlinear regression in Eq. (2) were manually determined. Particularly, to obtain reasonable fitting results, initial atomic coordinates of O_2 molecules were required to be near residues showing large ΔR_1 values. To abolish the initial parameter dependence of regression, parameters in Eq. (2) were estimated using the AIC¹⁹ and grid-searching (see Materials and Methods).

The AIC values with fitting results for one-, two-, and three-binding-site models were listed in Table S1. The three-binding-site model giving the minimum AIC value was considered to be the best possible solution. Consideration of the one-, two-, three-, and four-binding-site models is shown in Supplementary Material. The obtained parameters, C_1 , C_2 , C_3 , and C_0 were $2.6 \times 10^5 \text{ (\AA}^6/\text{s)}$, $2.8 \times 10^4 \text{ (\AA}^6/\text{s)}$, $4.6 \times 10^2 \text{ (\AA}^6/\text{s)}$, and $1.1 \text{ (s}^{-1}\text{)}$, respectively. The two O_2 -binding sites of WT* in solution closely match the locations of hydrophobic cavities 3 (site 2) and 4 (site 1) in the crystal structure [Fig. 3(C)]. Although the third binding site was identified at the molecular surface near residue 19 (site 3), O_2 association with site 3 is two-to-three orders of magnitude weaker than sites 1 and 2. Therefore, we do not consider this further.

As O_2 may transit between different internal binding locations on a time scale that is much faster than that associated with solvent exchange, the

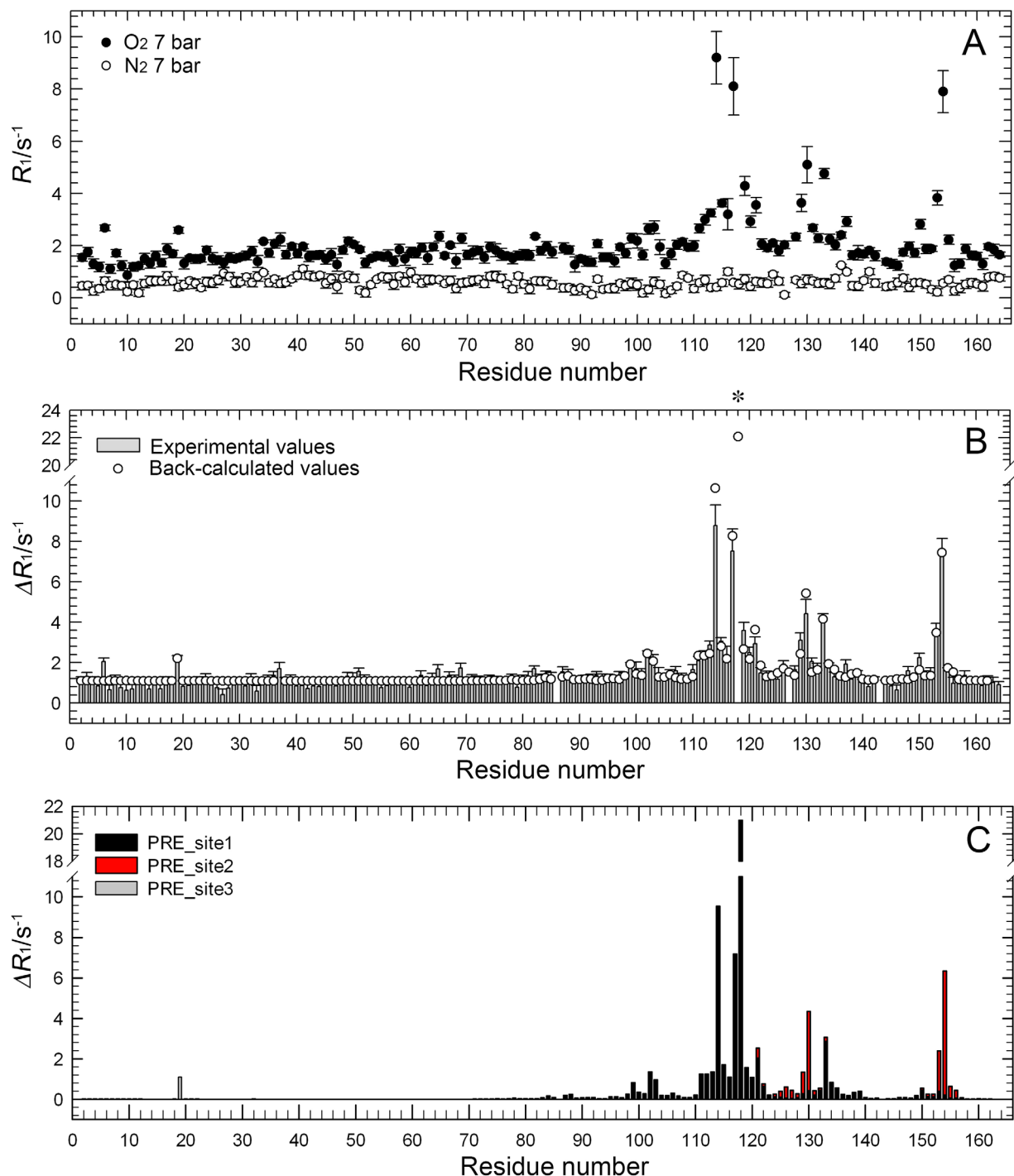


Figure 2. ^1H longitudinal relaxation enhancements for amide protons. (A) ^1H longitudinal relaxation rate constants, R_1 , for amide protons against WT* residue number at 7 bar of O_2 and N_2 . (B) Experimental and back-calculated O_2 -induced ^1H longitudinal relaxation enhancements, ΔR_1 , defined as the difference of R_1 between 0 mM (N_2 , 7 bar) and 8.9 mM O_2 . Severe line broadening at 7 bar of O_2 prohibited quantitative evaluation of ΔR_1 for Leu118 (asterisk). Residues 86 and 143 are proline. Assignment of Asp 127 is ambiguous. (C) Contributions of each O_2 -binding site to the back-calculated ΔR_1 .

parameters C_1 and C_2 typically reflect the average occupancy of O_2 in each cavity. However, it is noteworthy that the coefficients, C_i , can also be affected by the heterogeneity of the length-distribution, and the presence of internal motions of a PRE interaction vector (e.g., a O_2 - ^1H vector),^{20,21} which may affect the occupancy values to some extent. The

current results indicate about 10-times higher O_2 occupancy of cavity 4 (48 \AA^3) compared to that of cavity 3 ($14\text{--}15 \text{ \AA}^3$).¹⁰ As a point of reference, the cavity-enlarged T4 lysozyme mutant L99A (cavity 3: 25 \AA^3 and cavity 4: 150 \AA^3), was estimated to have 20-times higher occupancy of O_2 in cavity 4 compared to cavity 3.¹² However, these O_2 associations

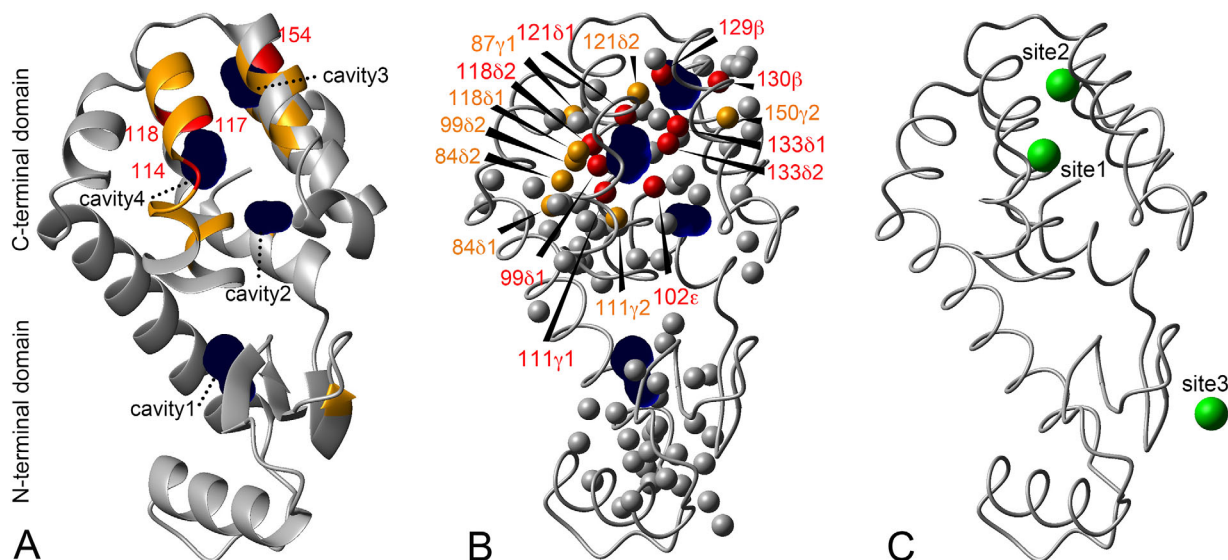


Figure 3. Mapping of O₂ binding sites. (A) Mapping of amide groups showing large ¹H longitudinal relaxation enhancements ($2.0 \text{ s}^{-1} < \Delta R_1 < 7.0 \text{ s}^{-1}$, orange; $\Delta R_1 \geq 7.0 \text{ s}^{-1}$ or remarkable signal loss, red) at 8.9 mM O₂ in WT*. Cavities in T4 lysozyme (PDB ID: 1L63), calculated by the program MOLMOL,³⁴ are depicted by dark-blue spheres (probe size 1.4 Å). (B) Mapping of methyl groups showing large ¹H longitudinal relaxation enhancements ($4.0 \text{ s}^{-1} < \Delta R_1 < 10 \text{ s}^{-1}$, orange; $\Delta R_1 \geq 10 \text{ s}^{-1}$ or remarkable signal loss, red) at 8.9 mM O₂ in WT*. (C) Potential O₂-binding sites (green spheres) in WT* as determined by ¹H longitudinal relaxation enhancements of amide protons. All images were prepared using MOLMOL.³⁴

with cavity 3 (site 2) are surprising, because the molecular volume of O₂ (53 Å³), which is derived from the van der Waals volume of a single oxygen atom, and the internuclear distance (1.21 Å), is much larger than the cavity volume. The rearrangement of the backbone and side-chain atoms might

have taken place around cavity 3. Note, however, that the chemical shift changes induced by O₂ in cavity 3 may not necessarily be observed, as the O₂ occupancy of cavity 3 is much lower than that of cavity 4. Back-calculated PRE contributions were plotted in Figure 2B, and the *R*² value of the

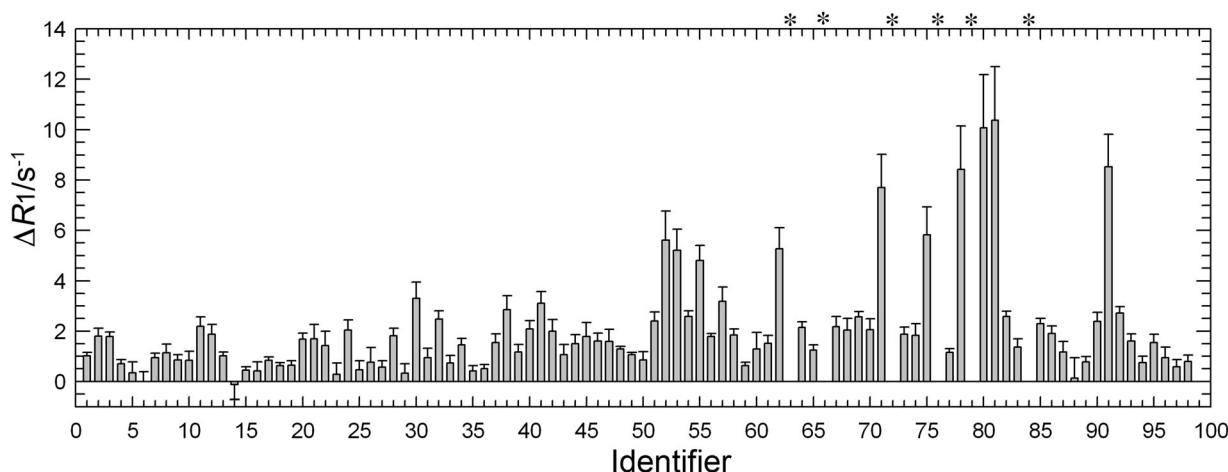


Figure 4. O₂-induced ¹H longitudinal relaxation enhancements, ΔR_1 , for methyl protons, defined as the difference of *R*₁ between 0 mM (N₂, 7 bar) and 8.9 mM O₂. Identifiers 1–98 have the following assignments: 1:M1ε, 2:I3γ2, 3:I3δ, 4:M6ε, 5:L7δ2, 6:L7δ1, 7:I9δ, 8:I9γ2, 9:L13δ1, 10:L15δ1, 11:L15δ2, 12:I17γ2, 13:I17δ, 14:T26γ, 15:I27δ, 16:I27γ2, 17:I29γ2, 18:I29δ, 19:L32δ1, 20:L32δ2, 21:L33δ2, 22:L33δ1, 23:T34γ, 24:L39δ2, 25:L39δ1, 26:A41β, 27:A42β, 28:L46δ1, 29:L46δ2, 30:A49β, 31:I50γ2, 32:I50δ, 33:T54γ, 34:V57γ1γ2, 35:I58γ2, 36:I58δ, 37:T59γ, 38:A63β, 39:L64δ2, 40:L66δ1, 41:L66δ2, 42:V71γ1, 43:V71γ2, 44:A73β, 45:A74β, 46:V75γ2, 47:V75γ1, 48:I78δ, 49:I78γ2, 50:L79δ2, 51:A82β, 52:L84δ1, 53:L84δ2, 54:V87γ2, 55:V87γ1, 56:V91δ2, 57:A93β, 58:V94γ2, 59:V94γ1, 60:A97β, 61:A98β, 62:L99δ2, 63:L99δ1, 64:I100δ, 65:I100γ2, 66:M102ε, 67:V103γ2, 68:V103γ1, 69:M106ε, 70:T109γ, 71:V111γ2, 72:V111γ1, 73:A112β, 74:T115γ, 75:L118δ1, 76:L118δ2, 77:M120ε, 78:L121δ2, 79:L121δ1, 80:A129β, 81:A130β, 82:V131γ2, 83:V131γ1, 84:L133δ2, 85:L133δ1, 86:A134β, 87:T142γ, 88:A146β, 89:V149γ2, 90:V149γ1, 91:I150γ2, 92:I150δ, 93:T151γ, 94:T152γ, 95:T155γ, 96:T157γ, 97:A160β, 98:L164δ1. Severe peak loss prohibited the quantitative evaluation of ΔR_1 for several methyl peaks (asterisks).

correlation between the back-calculated and observed ΔR_1 values was 0.97, indicating a high correlation. Moreover, the contribution of each binding site to ΔR_1 was calculated using Eq. (2) [Fig. 2(C)]. This analysis showed a quantitative contribution of each binding site to ΔR_1 . For instance, residues 124–130 and 153–156 exhibited dominant contributions from the cavity 3 binding site, although the O_2 binding probability in cavity 3 was 10 times smaller than that in cavity 4.

Background PREs, which correspond to parameter C_0 , can be explained by the contribution of O_2 dissolved in the solvent when no specific protein binding sites exist. In the diffusion model of dipole-dipole electron-nuclear interactions, the PRE contribution is inversely proportional to the cubic distance of the closest approach between nuclei and unpaired electron.^{14,22} PRE contribution for the surface protons was estimated to be approximately 1.6 s^{-1} at 8.9 mM O_2 , when assuming 2.72 Å as the closest approach. The estimate would be much smaller for protons buried in the interior of WT* (i.e., 0.26 s^{-1} at 5 Å as the closest approach). However, the observed site-specific difference of amide proton PREs from the O_2 in the solvent seems to be small. The standard-deviation of PREs from the average, which reflects the site-specific variation of PREs, is 0.30 s^{-1} for residues 2–90. This amount is similar to that of the experimental error of PREs (the average error, 0.22 s^{-1}). In contrast, a PRE contribution larger than those estimated cannot be explained without the specific binding of O_2 in the protein; the background PRE contribution was assumed to be constant for all amide protons to simplify the analysis. Indeed, the strong correlation between the back-calculated and observed ΔR_1 values indicates that this assumption remained reasonable.

Next, the observed ΔR_1 values for methyl protons in the C-terminal domain were globally fitted to Eq. (2). The parameters obtained from the analysis for amide protons were used as initial parameters for the regression. The back-calculated ΔR_1 values and potential O_2 -binding sites are shown in Figure S1(A). The potential O_2 -binding sites, as determined by ΔR_1 of the methyl protons of residues 71–160 [Fig. S1(B)], were similar, but not identical to those of the amide protons. The R^2 value of the correlation between the back-calculated and observed ΔR_1 values for the methyl protons was 0.61, which is significantly lower than that for the amide protons. Similar difficulty was encountered for the regression analysis of the methyl protons from the L99A mutant protein.¹²

In general, methyl protons are located in much closer proximity to O_2 molecules in the protein cavities. As the distance to O_2 is $<3 \text{ Å}$, ΔR_1 , which has a $1/r^6$ distance dependence, increases very steeply. For example, the back-calculated ΔR_1 values for the

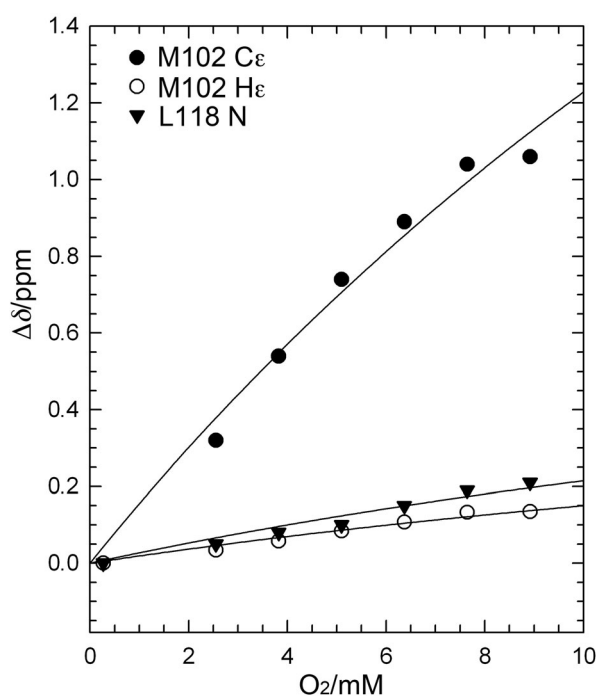


Figure 5. O_2 -induced chemical shift changes. Chemical shift changes of L118 amide nitrogen, M102 methyl protons, and M102 methyl carbon as a function of O_2 concentration.

methyl protons of L99 δ 1, M102 ϵ , V111 γ 1, and 118 δ 2 are 192, 30, 255, and 64 s^{-1} , respectively. While higher unambiguity of atomic positions is required for the analysis of methyl proton PREs, we obtained atomic coordinates of protons in the protein crystal using GROMACS energy minimization and an averaged atomic coordinate of three methyl protons for the analysis. However, such an ambiguity of atomic coordinates might affect the calculation of PREs. Accordingly, PREs of amide protons are considered more convenient and reliable for analyzing O_2 -binding sites in proteins.

Dissociation constants of O_2

Chemical shift changes of the L118 amide nitrogen, M102 methyl protons, and M102 methyl carbon are shown in Figure 5 as a function of O_2 concentration. All nuclei are located around cavity 4, and thus their chemical shift changes would be sensitive to O_2 occupancy. Similar O_2 -induced chemical shift changes were observed for both the L99A mutant of the protein¹² and OspA.¹⁶ As is likely to be the situation for both cases, the O_2 -induced chemical shift changes for the WT* were considered to originate from the paramagnetic shift, particularly the contact shift. As discussed above, O_2 molecules were found to associate to the hydrophobic cavities. Therefore, O_2 molecules may collide with nearby nuclei in the respective cavities, which can allow for delocalization of unpaired electrons to nuclei, thereby causing contact shift. This effect would be stronger for nuclei

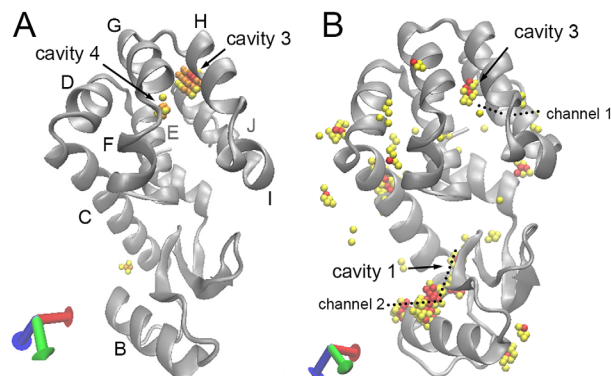


Figure 6. Average density maps of O₂ during MD simulations at 10 mM (A) and 100 mM (B) O₂. In panel (A), areas of 5–10, 10–100, or >100-times higher O₂ occupancy than the average are depicted by yellow, orange, or red spheres, respectively. In panel (B), areas of 5–10 and 10–50 times higher O₂ occupancy than the average are depicted by yellow or red spheres, respectively. Areas are depicted with 1 × 1 × 1 Å of mesh size. This image was prepared using the software VMD 2016.³⁵

having many electrons. If the pseudo-contact shift dominated, the magnitudes of chemical shift changes for the bound nuclei would be to a similar degree in ppm. However, similarity of O₂-induced chemical shifts among ¹H, ¹³C, and ¹⁵N nuclei were not observed. More details surrounding this observation were previously discussed.¹²

Assuming two-state exchange between the O₂-bound and -unbound forms of the protein, the association constant of O₂, *K*, can be estimated from the O₂ concentration dependence of chemical shifts, using the following Eq. (3):

$$\Delta\delta = \frac{\Delta\delta_{\max} \times K[\text{O}_2]}{1 + K[\text{O}_2]} \quad (3)$$

where [O₂] is the molar concentration of dissolved O₂ and Δδ_{max} is the nucleus-specific limiting value of the O₂-induced chemical shift. Data for three nuclei were globally fitted to Eq. (3) with a global variable, *K*. The association constant *K* was estimated to be 30 ± 10 M^{−1}, and the dissociation constant (*K_d* = 1/*K*) was 30 ± 10 mM. Therefore, 0.8% of the protein molecules were calculated to be in the O₂-bound state at atmospheric pressure (i.e., 0.27 mM O₂). Unfortunately, estimation of the association constant of O₂ to cavity 3 was difficult due to the small amplitude of O₂-induced chemical shifts. As discussed above, the relative occupancy of O₂ in cavity 3 as predicted by PREs was 10-times smaller than that in cavity 4, and thus the dissociation constant for the cavity 3 should be much higher than 30 mM.

O₂ binding and unbinding observed by MD simulations

Molecular dynamics (MD) simulations were performed in triplicate at 10 mM (corresponding to approximately 7 bar O₂ partial pressure) and 100 mM (corresponding to approximately 70 bar O₂

partial pressure) O₂ for 100 ns to investigate the interaction of O₂ with WT* (see Supplementary Material, Movies 1–3 for 10 mM O₂; Movies 4–6 for 100 mM O₂). When the system contains three O₂ molecules (10 mM O₂), one O₂ molecule was initially inserted in cavity 4 of WT* (PDB ID: 1L63) to study the dynamic aspects of O₂ in the hydrophobic cavities. Note that cavity 3 in the protein crystal structure, which is the initial structural model, cannot accommodate any O₂ molecules because the cavity space (15 Å³)¹⁰ was smaller than the molecular size of O₂. During the simulation, the inserted O₂ molecule frequently showed rotational motion as well as migration between cavities 3 and 4. Conformational fluctuation of the protein appears to allow association of O₂ with cavity 3. We also observed unbinding of O₂ from the protein through the cleft between helices G and H, while binding of O₂ to the protein was not observed at the O₂ concentration examined. The unbinding of O₂ was also observed in the cavity-enlarged L99A mutant protein, although O₂ molecules egressed through the cleft between helices D and G or H and J.¹² Figure 6(A) shows an O₂ density map calculated by 300 ns MD simulation with 10 mM O₂. In contrast to the PRE observations, O₂ occupancy in cavity 3 was much higher than that in cavity 4. Although the present MD simulations explained association of O₂ with the hydrophobic cavities 3 and 4, their O₂ occupancy seemed to be inconsistent with the PRE observations.

Reasons for the O₂ binding event not being observed in MD simulations at 10 mM O₂ are considered to be two-fold. First, the binding event is more rarer than the unbinding event. According to the *K_d* of O₂ (i.e., 30 mM), only 30% of the protein molecules were in the O₂ bound form at 10 mM O₂. Second, the exchange time scale might be much slower than the simulation time used in the present analysis (i.e., 300 ns). Because chemical shifts were changed as a function of O₂ concentration, the

exchange rate, which is the sum of the binding and unbinding rates, would be faster than the chemical shift difference between the bound and unbound forms, $\Delta\omega$. The estimated maximum $\Delta\omega$ in the present NMR measurements was 5 ± 2 ppm (4,700 rad s⁻¹) for the methyl carbon of M102 (data not shown). Therefore, the residence time of O₂ inside or outside the cavity would be shorter than 210 μ s. Even if the residence time is 1 μ s, much longer simulation times might be required to detect the O₂ binding event.

In order to observe O₂ binding more effectively, MD simulations were performed with 30 O₂ molecules (100 mM of O₂) in the system. No O₂ molecules were initially inserted in the internal cavities. O₂ binding events were observed at cavity 3 and the water-containing cavity 1 of the protein during the 300 ns simulation (see Supplementary Material, Movies 4–6). We observed O₂ binding into cavity 3 through the cleft between helices H and J. Although the O₂ molecule occasionally moved around cavities 3 and 4, it remained almost exclusively around cavity 3. It then egressed from the protein through the same cleft. The cleft may thus represent binding and unbinding pathways for O₂ molecules. The pathway to cavity 3 is designated channel 1, as indicated in Figure 6(B). The pathway was also observed in the MD simulation for the L99A protein mutant. In addition, we frequently observed association of O₂ with the water-containing cavity 1. Figure 6(B) shows an average density map of O₂ during 300 ns MD simulations with 100 mM O₂. More than 10-times higher O₂ occupancy than the average was observed at cavities 1 and 3 as well as on several regions of the protein's surface. In contrast to the PRE observations, O₂ frequently sampled voids in the N-terminal domain, particularly around cavity 1. As shown in Figure 6(B), O₂ molecules bound to and unbound from cavity 1 through the channel near the B-helix (i.e., channel 2). Although the binding of O₂ to cavity 1 was not shown by both experimental methods and through simulation using 10 mM O₂, the MD simulations suggest that cavity 1 has the potential to associate with O₂ at higher O₂ concentrations.

O₂ accessibility of internal cavities

The accessibility of O₂ to the internal cavities within three proteins, namely OspA,¹⁶ WT* T4 lysozyme (the present study), and its L99A mutant,¹² have been investigated by our groups using O₂ gas-pressure NMR spectroscopy. Because of the paramagnetic property and small size of O₂, ligand accessible sites, including novel binding sites, were detected in the aforementioned proteins. O₂ was in fact found to share binding sites with noble gases such as xenon, krypton, and argon.¹⁰ In addition, O₂ bound more specifically to hydrophobic rather than water-containing hydrophilic cavities. According to

the protein crystal structure, there are no channels to access internal cavities. Therefore, conformational fluctuation would be required for O₂ to bind to the internal cavities. MD simulation results qualitatively explained the NMR observations and suggested that conformational fluctuation takes place within the ground state ensemble to transiently develop channels that enable O₂ binding to and unbinding from internal cavities. Because O₂ shares binding sites with noble gases, benzene, and substituted benzenes in WT* and its L99A mutant,^{1,6,7,10} O₂ accessible cavities and pockets in this protein are potential binding sites for small hydrophobic molecules and fragments of compounds. More generally, ligand screening and the design of small molecules targeting O₂ accessible cavities and pockets might offer a rational strategy for controlling protein functional activity and stability.

Materials and Methods

Sample preparation

Uniformly ¹⁵N/¹³C-labeled cysteine-free T4 lysozyme (WT*, C54T/C97A) was produced in M9 media with ¹⁵NH₄Cl and ¹³C₆ glucose as the sole respective nitrogen and carbon sources, following established protocols.²³ The purified protein sample was dialyzed in a 50 mM phosphate buffer containing 25 mM NaCl at pH 5.5. The final concentration of the ¹⁵N/¹³C-labeled sample was 1.3 mM. Sample concentration was measured by UV absorption at 280 nm and was calculated with a molar extinction coefficient of 25,440 M⁻¹·cm⁻¹.

NMR experiments

We used a ¹H 600 MHz (Bruker BioSpin Co. AVANCE: Bruker, Billerica, MA) NMR spectrometer. The pressure resistance NMR tube (528-QPV-7, Wilmad-Lab Glass Co., Vineland, NJ) was connected to a gas cylinder by PTFE tubing and was used to study the binding of O₂ and nitrogen (N₂) to the protein. Mole fraction solubility of O₂ and N₂ in water at 298 K are 2.3×10^{-5} and 1.2×10^{-5} , respectively.²⁴ Applying gas pressure increases the concentrations of O₂ and N₂ dissolved in solvent. In this report, we used absolute pressure (gauge pressure + atmospheric pressure). ¹H-NMR, ¹H/¹⁵N HSQC, and ¹H/¹³C constant time (CT) HSQC spectra were obtained for purified protein at 298 K with different gas pressures. ¹H, ¹³C, and ¹⁵N chemical shift assignments of the protein were previously reported by Fischer et al.²⁵ ¹H longitudinal relaxation enhancements for amide and methyl protons were obtained from ¹H/¹⁵N HSQC and ¹H/¹³C CT-HSQC spectra using saturation recovery. This was achieved with proton x and y purge pulses followed by a relaxation delay before each scan. Ten relaxation delays ranging from 0.003 to 1.5 s were used.

Spectral analysis was performed using TopSpin (Bruker, Billerica, MA), NMRPipe,²⁶ and NMRViewJ.²⁷

PRE analysis and procedure for statistical model identification

PRE data were globally fitted to Eq. (2). To evaluate the possible solutions, we first introduce a score function $S(\vec{\theta})$, as given in Eq. (4):

$$S(\vec{\theta}) = \sum_{a=1}^m \frac{(\Delta R_1(a, \vec{\theta}) - \mu_{\text{obs},a})^2}{\sigma_{\text{obs},a}^2} \quad (4)$$

In the above equation, m is the number of residues, $\vec{\theta}$ is a set of all parameters in Eq. (2) (i.e., $C_0, X_1, Y_1, Z_1, C_1, X_2, Y_2, Z_2, C_2, \dots, X_n, Y_n, Z_n, C_n$), and $\mu_{\text{obs},a}$ and $\sigma_{\text{obs},a}$ are the PRE value, and its standard deviation for the amide proton of residue number a , respectively. $\Delta R_1(a, \vec{\theta})$ is the PRE of residue number a , which is estimated from the parameters $\vec{\theta}$ according to Eq. (2). Clearly, $S(\vec{\theta})$ is the squared sum of the differences between the predicted values $\Delta R_1(a, \vec{\theta})$ and the experimental values $\mu_{\text{obs},a}$, weighted according to the standard deviations. Thus, the value of $S(\vec{\theta})$ in Eq. (4) decreases when the predicted ΔR_1 values are better fitted to its experimental counterparts.

However, minimizing the value of $S(\vec{\theta})$ in Eq. (4) directly, by adding a number of binding sites, n , does not provide a physically meaningful result. This is because $S(\vec{\theta})$ can be decreased to zero by infinitely adding binding sites. Such a solution merely overfits the current set of experimental observations. Instead of choosing the smallest $S(\vec{\theta})$, we used the Akaike Information Criterion (AIC)¹⁹ to choose an appropriate binding model. The AIC is one of the most popular indices in statistics used for facilitating statistical model selection. The AIC value is given by Eq. (5):

$$\text{AIC} = 2k + \sum_{a=1}^m \log(2\pi\sigma_{\text{obs},a}^2) + S(\vec{\theta}) \quad (5)$$

where $k=4n+1$, which is the number of free parameters. The derivation of Eq. (5) is given in Supplementary Material. It has been shown that exploring the minimum AIC value corresponds to selecting the most predictive model (i.e., the best oxygen placement) to explain the observations (i.e., PREs) under some approximations.¹⁹ The AIC value consists of a penalty term (the first term), a constant term (the second term), and a score function (the third term). It increases when the number of binding sites increases, and decreases when the predicted ΔR_1 values are better fitted to its corresponding experimental values. The balance between the penalty and

the score prevents an oversimplified model (i.e., one with too small an n value, and with a bad prediction) or an overfitted model (i.e., one with too large an n value) from being selected.

Technically, minimizing the AIC value in Eq. (5) is difficult due to a variable number of sites, and multiple local minima in the function. We exploited the fact that when both the number of binding sites and the coordinates are fixed, the minimization problem simplifies to the well-known linear least square problem. In such a case, the AIC is described by Eqs. (6) and (7),

$$\text{AIC} = \text{const.} + \sum_{a=1}^m \left(\vec{Q}_a \cdot \vec{C} - \frac{\mu_{\text{obs},a}}{\sigma_{\text{obs},a}} \right)^2 \quad (6)$$

$$Q_{a,i} = \begin{cases} \sigma_{\text{obs},a}^{-1} & (i=0) \\ \sigma_{\text{obs},a}^{-1} \left((x_a - X_i)^2 + (y_a - Y_i)^2 + (z_a - Z_i)^2 \right)^{-3} & (\text{otherwise}) \end{cases} \quad (7)$$

In Eq. (6), \vec{Q}_a is an $n+1$ -dimensional vector, consisting of $Q_{a,0}, Q_{a,1}, \dots, Q_{a,n}$, and \vec{C} is also an $n+1$ -dimensional vector, consisting of C_0, C_1, \dots, C_n . Derivations of Eqs. (6) and (7) are presented in the Supplementary Materials file. Note that in Eq. (7), all variables to the right of $Q_{a,i}$ are either known or fixed, in this situation. Assuming that $C_i \geq 0$ ($i=0, 1, \dots, n$), minimizing the value of $Q_{a,i}$ in Eq. (7) is the same as solving the standard linear least-square problem with non-negative constraints. Because this optimization can be achieved easily, it is sufficient to optimize only the number of binding sites, and the coordinates of the O₂-binding sites to obtain the best possible solution.

A combination of grid-search and local refinement procedures was used to determine the best possible solution. First, the coordinate space was expressed in a $25 \times 29 \times 32$ grid (i.e., 23,200 grid points) with a 2-Å grid-space, each of which was a potential O₂-binding site. At each grid point, parameters C_0 and C_1 were optimized to minimize the AIC value in Eq. (6). Grid points showing the 100 smallest AIC values were selected for further optimization. Second, the 100 atomic coordinates of the potential O₂-binding site were optimized to minimize the AIC value using the covariance matrix adaptation evolution strategy (CMA-ES),²⁸ a derivative-free optimization method. In this CMA-ES, local optimum solutions were explored from among the 100 initial solutions. The best 10 non-overlapping coordinates selected within the 100 optimized coordinates were nominated for the first binding site. Note that optimized coordinates within 2 Å were considered as overlapping, and the coordinate showing the lowest AIC was retained for subsequent analysis. Third, the second binding site was explored within 23,200

grids assuming 10 different first binding sites (i.e., the nominated 10 binding sites). For the 232,000 pairs of the first and second binding sites, parameters in Eq. (2) (i.e., C_0 , C_1 , and C_2) were optimized to minimize the AIC. Next, the best 100 atomic coordinates of the second potential O₂-binding site were optimized using the CMA-ES. To exclude physically irrelevant solutions, we used a constraint that oxygen binding sites were not within r_{OH} (=2.72 Å) from all main-chain amide protons, which is the sum of van der Waals radii of oxygen and hydrogen. We further excluded solutions of which the oxygen coordinates converged within 0.01 Å from the edge of the coordinate space in the grid search. The edge was at least 10 Å apart from the protein surface. The best 10 non-overlapping coordinates were selected after the CMA-ES optimization. This procedure was repeated for up to a 4-binding site model. There were no valid solutions for the 4-binding site model, and thus further binding site models were not considered. The binding model and atomic coordinates with parameters giving the minimum AIC value were considered the best possible solution.

Molecular dynamics simulation

Molecular dynamics (MD) simulations of 100 nanoseconds were performed in triplicate at O₂ concentrations of 10 mM and 100 mM using the GROMACS 2016 molecular simulation package.²⁹ The system contained a WT* (PDB ID; 1L63), 3 (or 30) O₂ molecules, 8 chloride ions, and approximately 15,000 water molecules at a pressure of 7 bar (or 70 bar). In the case at 10 mM O₂, one O₂ molecule was initially inserted in the largest hydrophobic cavity (i.e., cavity 4). Both N- and C-termini of WT* were uncapped (i.e., charged). The OPLSL-AA force field³⁰ was used for the protein, and the TIP4P model was used for water.³¹ Potential parameters for O₂ and chloride ions were as per those previously described in the literature.^{32,33} MD simulations were conducted with the *NPT* ensemble (300 K, 7 bar or 70 bar) in a truncated dodecahedron box with dimensions of 88.7 Å. In addition, to estimate the rotational correlation times of O₂ in cavity 3, we performed separate one nanosecond MD simulations. Details of MD simulations were previously described.¹²

Acknowledgments

We thank Dr. Yuichi Yoshimura for providing the NMR pulse sequence and Mr. Takuro Wakamoto for kindly helping with the NMR spectral analysis.

References

1. Morton A, Matthews BW (1995) Specificity of ligand-binding in a buried nonpolar cavity of T4 lysozyme:

- linkage of dynamics and structural plasticity. *Biochemistry* 34:8576–8588.
2. Otting G, Liepinsh E, Halle B, Frey U (1997) NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. *Nat Struct Biol* 4:396–404.
3. Hubbard SJ, Argos P (1996) A functional role for protein cavities in domain:domain motions. *J Mol Biol* 261:289–300.
4. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat Struct Biol* 3:178–187.
5. Mulder FAA, Mittermaier A, Hon B, Dahlquist FW, Kay LE (2001) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* 8:932–935.
6. Nucci NV, Fuglestad B, Athanasoula EA, Wand AJ (2014) Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc Natl Acad Sci Usa* 111:13846–13851.
7. Eriksson AE, Baase WA, Wozniak JA, Matthews BW (1992) A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* 355:371–373.
8. Morton A, Baase WA, Matthews BW (1995) Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* 34:8564–8575.
9. Nicholson H, Anderson DE, Dao-pin S, Matthews BW (1991) Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* 30:9816–9828.
10. Quillin ML, Breyer WA, Griswold IJ, Matthews BW (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J Mol Biol* 302:955–977.
11. Llinas M, Gillespie B, Dahlquist FW, Marqusee S (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* 6:1072–1078.
12. Kitahara R, Yoshimura Y, Xue MJ, Kameda T, Mulder FAA (2016) Detecting O₂ binding sites in protein cavities. *Sci Rep* 6:20534.
13. Bezsonova I, Forman-Kay J, Prosser RS (2008) Molecular oxygen as a paramagnetic NMR probe of protein solvent exposure and topology. *Concept Magn Reson A* 32A:239–253.
14. Teng CL, Bryant RG (2004) Mapping oxygen accessibility to ribonuclease a using high-resolution NMR relaxation spectroscopy. *Biophys J* 86:1713–1725.
15. McNaughton L, Hernandez G, LeMaster DM (2003) Equilibrium O₂ distribution in the Zn²⁺-protoporphyrin IX deoxymyoglobin mimic: application to oxygen migration pathway analysis. *J Am Chem Soc* 125:3813–3820.
16. Kawamura T, Wakamoto T, Kitazawa S, Sakuraba S, Kameda T, Kitahara R (2017) Analysis of O₂-binding sites in proteins using gas-pressure NMR spectroscopy: outer surface protein A. *Biophys J* 112:1820–1828.
17. Sakakura M, Noba S, Luchette PA, Shimada I, Prosser RS (2005) An NMR method for the determination of protein-binding interfaces using dioxygen-induced spin-lattice relaxation enhancement. *J Am Chem Soc* 127:5826–5832.
18. Hernandez G, Teng CL, Bryant RG, LeMaster DM (2002) O₂ penetration and proton burial depth in proteins: applicability to fold family recognition. *J Am Chem Soc* 124:4463–4472.
19. Akaike H (1974) New look at statistical-model identification. *IEEE Trans Automat Contr* 19:716–723.

20. Iwahara J, Schwieters CD, Clore GM (2004) Ensemble approach for NMR structure refinement against ^1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 126:5879–5896.
21. Iwahara J, Clore GM (2010) Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long, variable-length vectors using NMR paramagnetic relaxation enhancement. *J Am Chem Soc* 132:13346–13356.
22. Ulmer TS, Campbell ID, Boyd J (2002) The effects of dissolved oxygen upon amide proton relaxation and chemical shift in a perdeuterated protein. *J Magn Reson* 157:181–189.
23. Mulder FAA, Hon B, Muhandiram DR, Dahlquist FW, Kay LE (2000) Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multi-nuclear NMR. *Biochemistry* 39:12614–12622.
24. Scharlin P, Battino R, Silla E, Tunon I, Pascual-Ahuir JL (1998) Solubility of gases in water: correlation between solubility and the number of water molecules in the first solvation shell. *Pure Appl Chem* 70:1895–1904.
25. Fischer MW, Majumdar A, Dahlquist FW, Zuiderweg ER (1995) ^{15}N , ^{13}C , and ^1H NMR assignments and secondary structure for T4-lysozyme. *J Magn Reson B* 108:143–154.
26. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6:277–293.
27. Johnson BA, Blevins RA (1994) NMR View: a computer-program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614.
28. Hansen NA, Ostermeier A (1996) Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In: *Proceedings of the 1996 IEEE Intern. Conf. on Evolutionary Computation*, pp 312–317.
29. Berendsen HJC, van der Spoel D, van Drunen R (1995) Gromacs: a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun* 91:43–56.
30. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236.
31. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
32. Chandrasekhar J, Spellmeyer DC, Jorgensen WL (1984) Energy component analysis for dilute aqueous-solutions of Li^+ , Na^+ , F^- , and Cl^- ions. *J Am Chem Soc* 106:903–910.
33. Arora G, Sandler SI (2006) Mass transport of O_2 and N_2 in nanoporous carbon (C168 Schwarzite) using quantum mechanical force field and molecular dynamics simulations. *Langmuir* 22:4620–4628.
34. Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55.
35. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38.